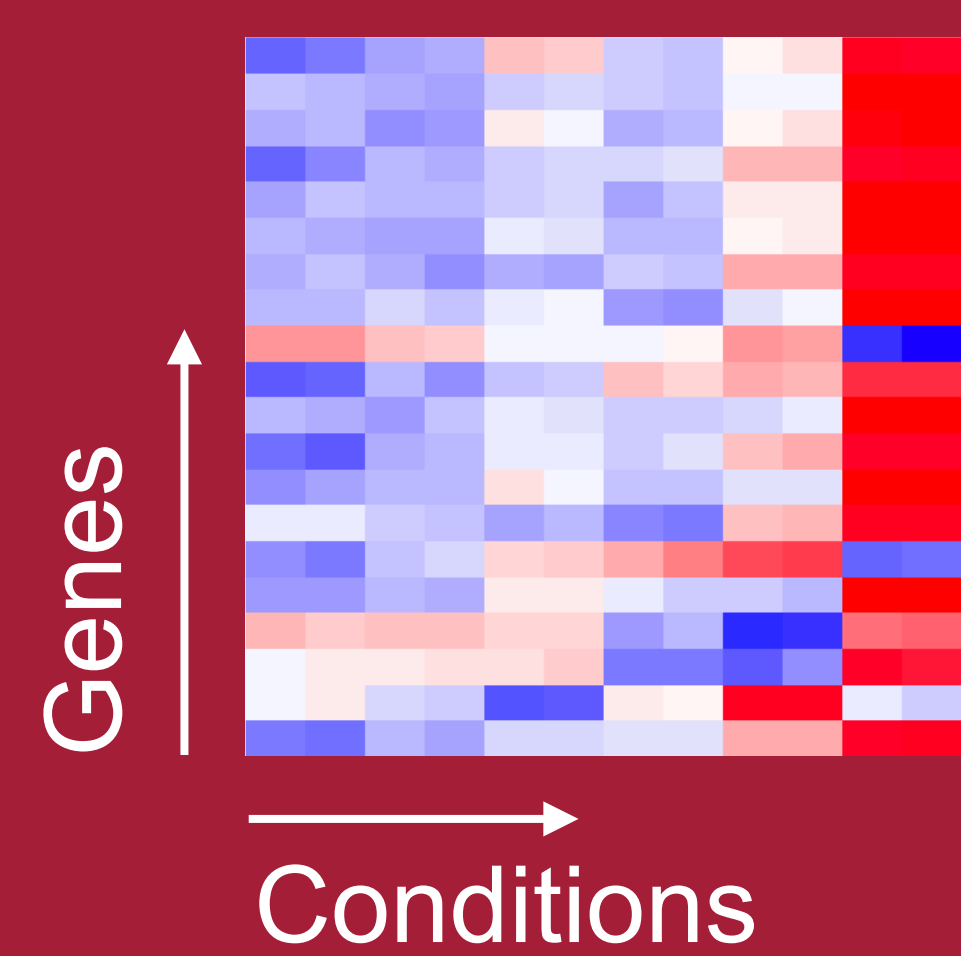


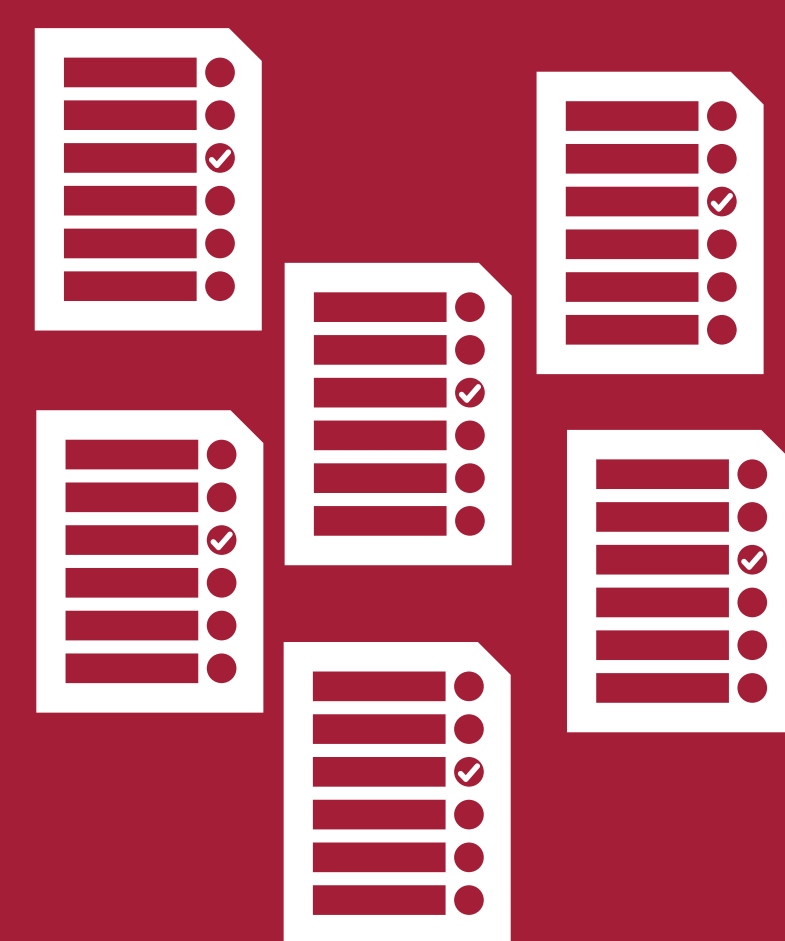


Differentially expressed genes in multiple conditions

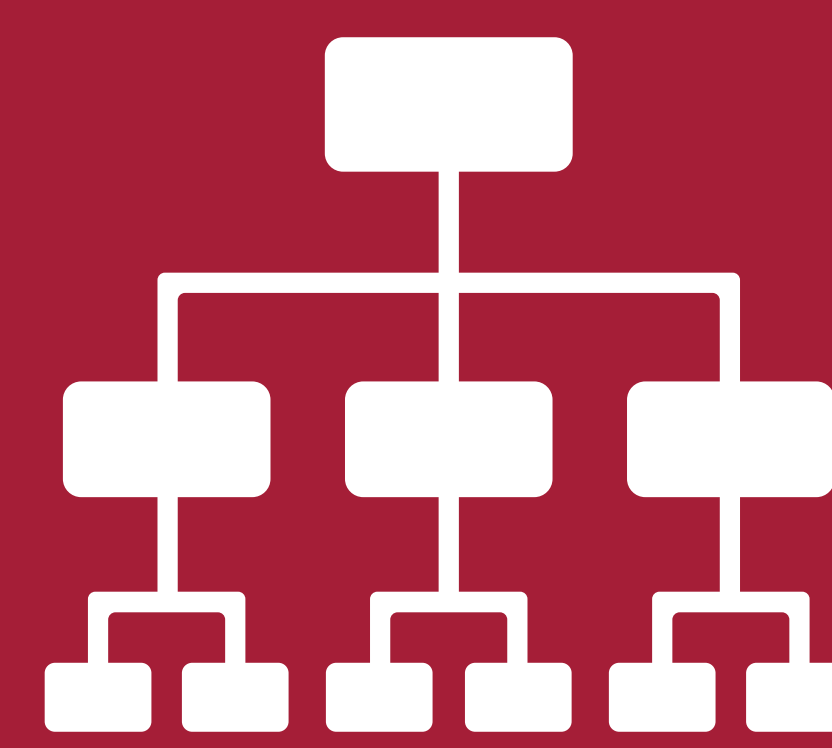


GO enrichment

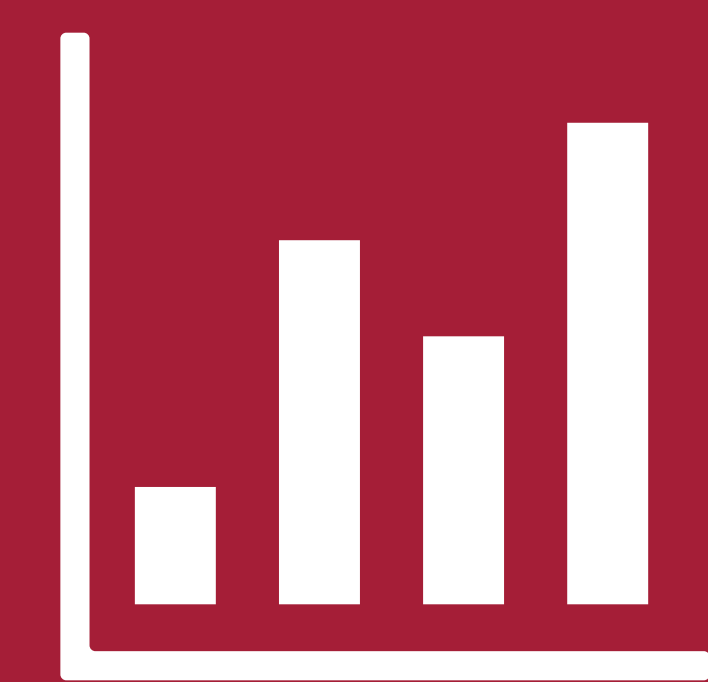
GO lists



Redundancy clustering of GO Terms



Comparative Visualization



Background

Analysis pipelines for RNA sequencing data often produce long lists of differentially expressed genes, for example in time series experiments or when comparing modified strains. For functional analysis, these lists can be analyzed using Gene Ontology (GO) enrichment analysis resulting in lists of enriched GO terms with associated p-values.

Overview

Since GO terms are organized roughly hierarchical, an enrichment of a child term propagates to enrichments of parent terms which can lead to redundancy in the list. Moreover, GO lists can be very long and hard to compare. Therefore we developed a web application that first clusters terms hierarchically based on their dispensability. In a visualization dashboard users can filter this clustering and compare the resulting lists.

Algorithm

An adapted version of the REVIGO algorithm [1] is applied to the lists of GO terms for redundancy clustering.

For every pair of terms semantic similarity is computed and the pairs are sorted. Starting with the most similar pair, a term is rejected if one of multiple criteria is fulfilled. The resulting structure can be represented as a tree where rejected terms are child nodes of accepted terms.

The rejection criteria are (applied in this order):

1. The term has a **frequency** in the background data $> 5\%$
2. The majority of **p-values** is less significant
3. The term is the **parent** of the other term and contains more than 75% of its genes
4. The term is a **child** of the other term
5. (Pseudo) random rejection

Visualization

On this poster we show a time series dataset with nine conditions, or more precisely timepoints [2].

The clustering is visualized as a clustered heatmap **a** where the heatmap shows the significance of the GO terms at the different conditions.

With the right slider redundant terms can be filtered out. The left slider cuts the tree at a specific dispensability to produce flat clusters.

A PCA plot of the lists as well as a correlation heatmap summarize the data **b c**.

Based on the flat clusters the GO list of every condition is visualized as a treemap in a small multiple plot where the size of a rectangle corresponds to the negative logarithm of the p-value **f**. Users can select a treemap to visualize it in the main treemap view where the labels of the GO terms are shown **e**.

In order to show the change between conditions, switching treemaps triggers an animation where the rectangles change their size dynamically.

For a more detailed comparison single GO terms can be selected via a hover interaction and their p-values are visualized in a simple bar chart or line chart depending on the data **d**.

Future work

We plan to extend our tool to be able to compare lists of GO terms of different species, to evaluate it with potential users, and to make it available to a broad audience.

References

[1] Supek, Fran, et al. "REVIGO summarizes and visualizes long lists of gene ontology terms." *PloS one* 6.7 (2011): e21800.

[2] Sulheim, Snorre, et al. "Enzyme-constrained models and omics analysis of *Streptomyces coelicolor* reveal metabolic changes that enhance heterologous production." *iScience* (2020): 101525.